

• 论 著 •

基于神经网络模型和血常规指标的孕妇地中海贫血预测研究*

龙尧水, 白文学[△]

重庆市江津区妇幼保健院检验科, 重庆 402260

摘要:目的 运用神经网络模型, 研究重庆市江津区地中海贫血(简称地贫, 即珠蛋白生成障碍性贫血)孕妇的血常规指标对地贫的预测效果。方法 收集 2016 年 1 月至 2022 年 12 月该院 9 652 例产前检查孕妇的临床资料、血常规及地贫基因检测结果。在 9 652 例孕妇中, 检出地贫基因阳性 847 例, 阴性 8 805 例。在地贫基因阳性和阴性孕妇中分别随机抽取 424 例和 4 402 例作为试验组, 剩余 423 例地贫孕妇与 4 403 例阴性孕妇作为对照组。构建随机森林模型、Logistic 回归模型分析地贫预测效果较好的血常规指标, 建立神经网络模型区分地贫类型为 α 或 β 。结果 847 例地贫阳性孕妇中, α 地贫阳性 569 例 (5.90%), β 地贫阳性 267 例 (2.77%), α 合并 β 地贫阳性 11 例 (0.11%)。通过随机森林模型筛选出对诊断地贫重要性较高的 5 项血常规指标平均红细胞体积(MCV)、平均血红蛋白含量(MCH)、红细胞体积分布宽度变异系数(RDW-CV)、平均血红蛋白浓度(MCHC)、红细胞计数(RBC), 采用 Logistic 回归模型进行验证, 受试者工作特征(ROC)曲线分析显示, 5 项指标的曲线下面积(AUC)分别为 0.906、0.904、0.785、0.783、0.780, 模型 AUC 为 0.906。采用随机森林筛选的血常规指标, 建立神经网络模型预测 α 地贫、 β 地贫并优化, 得出 α 、 β 地贫预测值均为 100%。结论 该研究建立的基于血常规指标的神经网络模型对 α 、 β 地贫预测率较高, 可为地贫的早期筛查提供了新的思路。

关键词: 地中海贫血; 血常规指标; 神经网络模型; 孕妇

DOI: 10.3969/j.issn.1673-4130.2023.20.003

中图法分类号: R446.11+1; R714.254

文章编号: 1673-4130(2023)20-2447-06

文献标志码: A

Prediction of thalassemia in pregnant women based on neural network
model and blood routine indexes*

LONG Yaoshui, BAI Wenxue[△]

Department of Clinical Laboratory, Maternal and Child Health Hospital of Jiangjin
District of Chongqing, Chongqing 402260, China

Abstract: **Objective** To investigate the prediction effect of blood routine indexes on thalassemia in pregnant women in Jiangjin district of Chongqing by using neural network model. **Methods** From January 2016 to December 2022, the clinical data, blood routine and thalassemia gene test results of 9 652 pregnant women were collected. Among 9 652 pregnant women, 847 were positive and 8 805 were negative for thalassemia gene. Among the positive and negative cases, 424 and 4 402 were randomly selected as the test group. In contrast, the remaining 423 thalassemia pregnant women and 4 403 negative pregnant women were assigned to the control group. The random forest model and Logistic regression model were constructed to analyze the blood routine indexes of thalassemia with good prediction effect, and the neural network model was established to differentiate between α -thalassemia and β -thalassemia. **Results** Among 847 pregnant women with thalassemia genes, 569 (5.90%) were positive for α -thalassemia, 267 cases (2.77%) were positive for β -thalassemia, and 11 (0.11%) were positive for α combined with β thalassemia. Furthermore, five blood routine indexes MCV, MCH, RDW-CV, MCHC and RBC, which were of high importance in the diagnosis of thalassemia, were selected by random forest model. These five indexes were verified by Logistic regression model. Receiver operating characteristic (ROC) curve analysis showed that the area under the curve (AUC) of the five indexes were 0.906, 0.904, 0.785, 0.783 and 0.780, respectively, and the overall AUC for the model was 0.906. Using the blood routine indexes of random forest screening, a neural network model was established to predict and optimize α and β thalassemia, and the predicted values of both α and β thalassemia were 100%. **Conclusion** The prediction rate of α and β thalassemia based on the neural network model established in this study is high, which provides a new idea for the early screening of thalassemia.

Key words: thalassemia; blood routine indices; neural network model; pregnant women

* 基金项目: 重庆市科卫联合项目(2022MSXM092)。

作者简介: 龙尧水, 男, 主管技师, 主要从事分子生物学研究。 [△] 通信作者, E-mail: 6680806@qq.com。

地中海贫血(简称地贫),即珠蛋白生成障碍性贫血,是全球范围内最普遍的单基因遗传病之一,为珠蛋白合成障碍导致的溶血性贫血^[1-2]。在南非、中东、东南亚一带高度流行,国内多发于南方沿海城市及西部农村等中低收入地区。中国拥有全球最多的地贫新发病例和患者,约 3 000 万中国人有地贫相关突变,300 万人患有中重型地贫,对家庭及社会造成沉重负担^[3-4]。由于地贫为常染色体隐性遗传病,故临床表现正常的父母也可能生下地贫婴儿,如果父母均为地贫基因携带者,极有可能生下重型地贫患儿,重症地贫存活率低且需要终身输血及输血治疗的各种并发症,故对地贫的早期诊断尤其重要^[5-6]。目前可用的地贫筛查和诊断方法包括血常规检测、渗透性脆性试验、目标红细胞(RBC)的血液涂片评估、包涵体鉴定、红细胞指数、血红蛋白(Hb)电泳、高效液相色谱(HPLC)和基因检测^[7-8]。地贫基因检测虽然作为诊断地贫的金标准,不过由于检测周期较长且费用较高,基层医院设备、人员缺乏等因素,地贫的群体筛查仍存在一定困难。而血常规检测是使用最广泛、最成熟的检验手段,如果将血常规指标作为地贫筛查指标,可以极大提升筛查效率并降低成本。

血常规指标检测是对血细胞数量和形态进行分析,通过血细胞分布和特定参数,对疾病进行辅助诊断,其中红细胞计数(RBC)、Hb、红细胞比容(HCT)、平均红细胞体积(MCV)、平均血红蛋白含量(MCH)、红细胞体积分布宽度(RDW)联合检测对地贫的辅助诊断参考价值受到广泛报道^[9-12]。Hb 和 RBC 在判断贫血方面起着决定性的作用。Hb 是红细胞的主要成分,将氧气输送到器官和组织。与红细胞相比,Hb 更能反映贫血的程度。MCV、MCH、RDW 与血液营养状况密切相关。国家地贫防控指南规定,MCV<80 fL 和(或)MCH<27 pg 为血常规筛查阳性。RDW 是血细胞分析仪通过快速测量大量红细胞的体积,得出红细胞体积的异质性。地贫患者 Hb、HCT、MCV、MCH、平均血红蛋白浓度(MCHC)、RDW 降低,符合地贫小细胞低色素性贫血的血常规指标分布特点^[13-14]。

本研究回顾性分析 2016—2022 年重庆江津地区地贫孕妇基因检测及血常规结果,并运用神经网络模型分析血常规相关指标对地贫的预测结果,旨在为本地区地贫孕妇的早期诊断、遗传咨询及治疗提供新的策略。

1 资料与方法

1.1 一般资料 收集 2016 年 1 月至 2022 年 12 月本院产前检查孕妇的病例资料、血常规及地贫基因检测结果。纳入标准:(1)首次到本院开展产前检查;(2)未合并其他感染性疾病。排除标准:(1)临床资料不全、重复;(2)近期使用引起红细胞增减的特殊药物。根据纳入排除标准,共收集病例 9 652 例。其中早孕期 8 029 例(83.2%),中孕期 1 417 例(14.7%),晚孕

期 206 例(2.1%)。9 652 例孕妇年龄 15~46 岁,平均年龄(26.99±4.85)岁。本研究经本院医学伦理委员会研究批准,所有受试者均签署知情同意书。

1.2 仪器与试剂 血常规指标检测采用日本 Sysmex XN1000 全自动血细胞分析仪及配套试剂。地贫基因检测采用杭州博日基因扩增仪 TC-96/G/H(b)和广东凯普医用核酸分子杂交仪 HB-2012A,以及 α -、 β -地贫基因检测试剂盒(PCR+膜杂交法)。

1.3 方法

1.3.1 血常规指标检测 采用真空采血管抽取研究对象静脉血 2 mL 于乙二胺四乙酸二钾(EDTA-K₂)抗凝管中,颠倒混匀 8 次,备用,血常规指标检测需在 4 h 内完成。采用 XN-1000 自动血细胞分析检测红细胞计数(RBC)、Hb、红细胞比容(HCT)、MCV、平均血红蛋白含量(MCH)、MCHC、红细胞体积分布宽度标准差(RDW-SD)、红细胞体积分布宽度变异系数(RDW-CV)等。

1.3.2 地贫基因检测 采用 PCR 结合导流杂交技术检测 3 种缺失型地贫($-\alpha^{SEA}$ 、 $-\alpha^{3.7}$ 和 $-\alpha^{4.2}$)、3 种突变类型 α 地贫(CS、QS 和 WS)和 17 种突变类型 β 地贫[-28(A-G)、-29(A-G)、-30(T-C)、-32(C-A)、CD14/15(+G)、CD17(A-T)、CD27/28(+C)、CD31(-C)、CD41/42(-TCTT)、CD43(G-T)、CD71/72(+A)、IVS-I-1(G-T, G-A)、IVS-I-5(G-C)、IVS-II-654(C-T)、 βE (G-A)、CAP(A-C, A-AAAC)、Int(T-G)]。

1.4 模型构建 在 9 652 例孕妇中检出地贫基因阳性 847 例,阴性 8 805 例,在阳性和阴性中分别随机抽取 424 和 4 402 例作为试验组,剩余的 423 例地贫基因阳性孕妇和 4 403 例阴性孕妇作为对照组。由于地贫阴性数据量较大,作为对照组与阳性数据对比会出现较大误差,数据平均分配可以确保建模过程受到阳性或阴性数据的干扰最小,防止过拟合现象出现,提高模型的泛化能力。采用随机森林(Random Forest)模型计算 RBC、Hb、HCT、MCV、MCH、MCHC、RDW-SD 等血常规指标的重要性,均方差增量(%IncMSE)、节点纯度增量(IncNodePurity)均表示变量的重要性,值越大重要性越高。建立随机森林模型评估各血常规指标的重要性,采用 Logistic 回归构建预测模型预测地贫发生的保护因素和危险因素,计算出各指标的优势比(OR);OR=1 表示无相关性,>1 为地贫发生的危险因素,<1 为地贫发生的保护因素。绘制受试者工作特征(ROC)曲线的验证模型的预测值,曲线下面积(AUC)越大表示预测效果越好。为了进一步区分地贫孕妇类型为 α 或 β ,构建了神经网络模型,569 例 α 地贫与 267 例 β 地贫中各随机 285 例 α 地贫与 134 例 β 地贫作为训练集,剩余数据作为验证集,分析模型对地贫类型的预测效果。随后选取最优的特征并进行优化,得到最终结果。

1.5 统计学处理 采用 Excel 2019 进行数据整理,

采用 R4.3.1 建立神经网络模型训练并验证。以 $P < 0.05$ 为差异有统计学意义。

2 结 果

2.1 地贫基因检测阳性情况 9 652 例病例中共检测出地贫阳性 847 例,阳性率为 8.78%,其中 α 地贫阳性 569 例,占比 5.90%; β 地贫阳性 267 例,占比 2.77%; α 及 β 合并基因型 11 例,占比 0.11%。与对

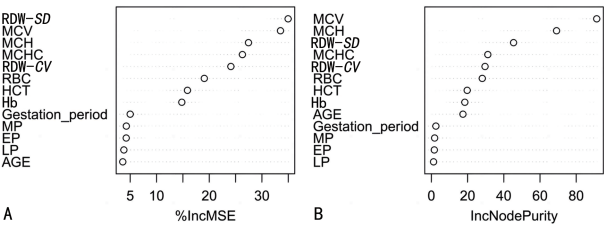
照组孕妇比较,地贫组血常规指标 Hb、MCV、MCH、MCHC、RDW-SD 降低,RBC、RDW-CV 增高,符合地贫小细胞低色素性贫血的临床表现,且 β 地贫、 α 合并 β 地贫基因型贫血改变幅度较 α 地贫变化更大。因孕妇就诊时主要为早孕期,与地贫关系不明确,但是随孕周增大,孕妇的贫血症状会加剧,故结果展示了不同孕周孕妇地贫基因检测结果。见表 1。

表 1 地贫基因检测情况[M(最小值,最大值)或 n(%)]

指标	α 地贫 (n=569)	β 地贫 (n=267)	α 合并 β 地贫 (n=11)	对照组 (n=8 805)	合计 (n=9 652)
RBC	4.53(2.79,6.84)	4.82(2.92,6.71)	4.94(3.76,6.07)	4.06(2.17,6.54)	4.10(2.17,6.84)
Hb	116(62,166)	101(67,164)	104(97,114)	121(48,177)	121(48,177)
HCT	35.9(21.0,48.4)	31.8(21.8,45.5)	32.3(31.4,35.1)	36.3(19.8,58.7)	36.1(19.8,58.7)
MCV	81.2(54.6,97.6)	65.2(55.7,99.1)	65.4(53.0,89.4)	89.9(53.9,122.0)	89.4(53.0,122.0)
MCH	26.3(17.1,33.6)	20.7(17.4,35.7)	21.4(16.5,28.7)	30.2(13.0,41.7)	30.0(13.0,41.7)
MCHC	323(272,364)	316(290,365)	321(309,328)	336(240,382)	335(240,382)
RDW-SD	39.2(32.4,83.7)	36.3(29.6,61.9)	34.8(31.1,53.8)	41.6(30.3,77.4)	41.4(29.6,83.7)
RDW-CV	13.9(11.1,31.3)	16.1(11.7,23.9)	15.0(13.1,22.6)	12.7(10.8,31.6)	12.8(10.8,31.6)
孕周					
早孕期	450(79.1)	203(76.0)	9(81.8)	7 367(83.7)	8 029(83.2)
中孕期	92(16.2)	48(18.0)	2(18.2)	1 275(14.5)	1 417(14.7)
晚孕期	27(4.7)	16(6.0)	0(0.0)	163(1.9)	206(2.1)

2.2 地贫危险因素分析 Logistic 回归模型分析显示,RBC、RDW-CV、晚孕期为地贫的危险因素,Hb、HCT、MCV、MCH、MCHC、RDW-SD、早孕期为地贫的保护因素($P < 0.001$),见表 2。

通过随机森林模型分析显示,重要性排序前 6 位依次为 RDW-SD、MCV、MCH、MCHC、RDW-CV、RBC,提示以上因素均对地贫的预测效果较好。基于 Logistic 回归模型,绘制 ROC 曲线分析血常规指标对地贫的预测结果,结果显示,MCV、MCH、RDW-CV、MCHC、RBC 等指标的 AUC 分别为 0.906、0.904、0.785、0.783、0.780,模型的 AUC 为 0.906,提示模型 AUC 与 MCV 结果一致。分析原因可能为 MCV 指标诊断地贫灵敏度较高。但当诊断的灵敏度提高后,相对应的误差也会增高,而模型整合了 MCH、RDW-CV、MCHC 等指标,预测地贫稳定性更高。见图 1。



注:A 为 %IncMSE;B 为 IncNodePurity。

图 1 随机森林模型试验组变量重要性分析

2.3 地贫危险因素的验证和地贫预测模型的建立 将对照组数据导入 Logistic 回归模型,绘制 ROC 曲线验证模型的预测价值,结果显示 MCH、MCV、MCHC、RDW-CV、RBC 等预测地贫的 AUC 分别为

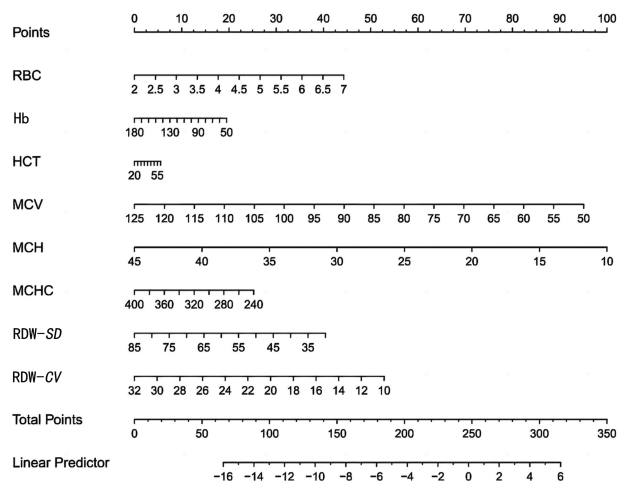
0.919、0.915、0.824、0.803、0.794,模型的 AUC 为 0.919,较试验组血常规指标预测值有所提高。对照组模型 AUC 与 MCH 一致,表示 MCH 预测地贫的效率较其他血常规指标更高。将血常规指标 RBC、Hb、MCV、MCH、MCHC、RDW-SD 等作为变量,建立地贫预测模型,同时绘制列线图,根据模型中 RBC、MCV、MCH 等各个变量对结局变量的贡献程度,给每个变量的取值水平进行赋分(Points),再将各个评分相加得到总评分(Total Points),最终通过总评分与地贫发生率之间的关系,计算出个体血常规指标对地贫的线性预测值(Linear Predictor 分数) >0 分表示发生地贫。见图 2。

表 2 地贫危险因素分析

项目	OR	95%CI	P
年龄	1.01	0.991~1.020	0.449
RBC	9.50	8.146~11.085	<0.001
Hb	0.95	0.942~0.952	<0.001
HCT	0.87	0.853~0.888	<0.001
MCV	0.81	0.803~0.821	<0.001
MCH	0.62	0.604~0.634	<0.001
MCHC	0.92	0.913~0.924	<0.001
RDW-SD	0.73	0.715~0.753	<0.001
RDW-CV	1.58	1.528~1.643	<0.001
早孕期	0.70	0.588~0.830	<0.001
中孕期	1.19	0.984~1.438	0.073
晚孕期	2.84	2.011~3.999	<0.001

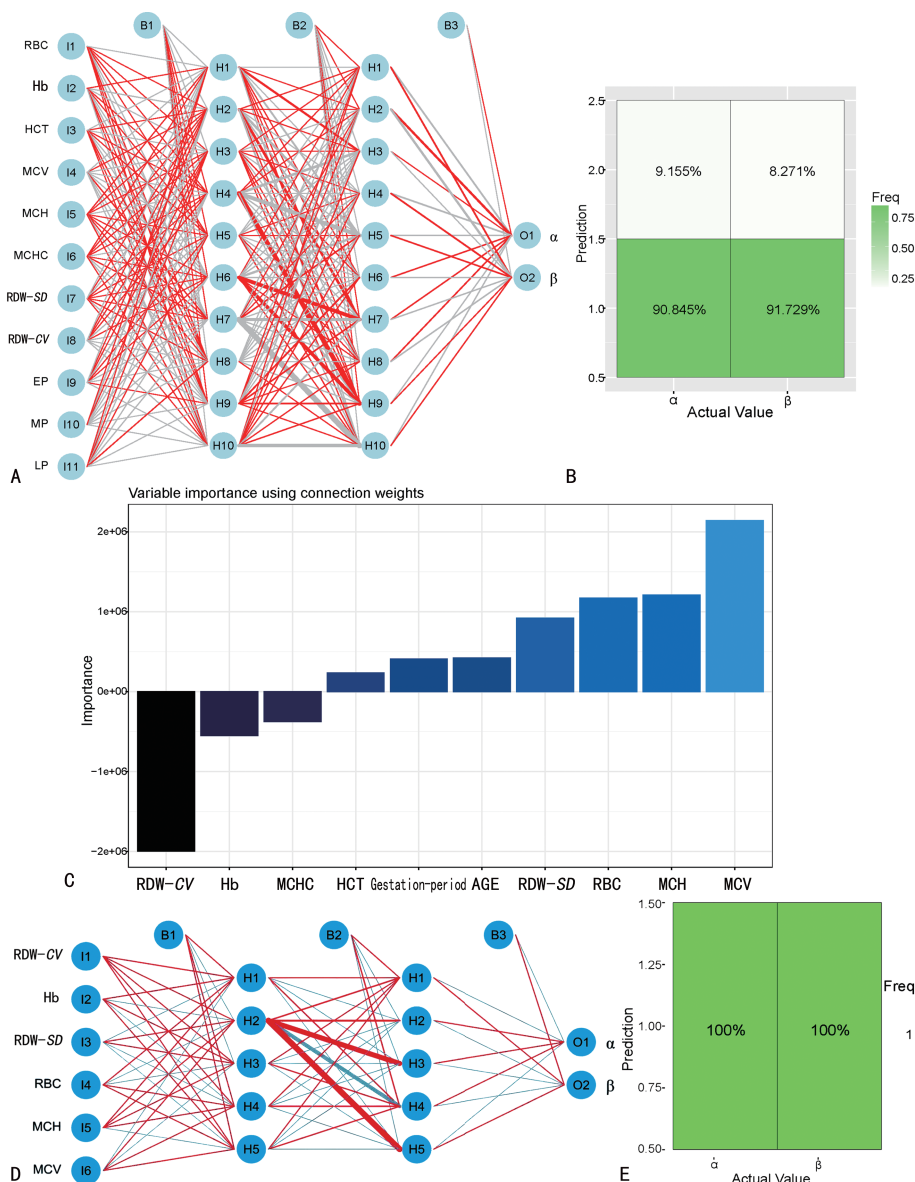
2.4 神经网络模型预测 α 、 β 地贫效果 为研究地贫孕妇的血常规指标、孕周等因素对地贫的预测效果,针对地贫患者血常规指标、孕周等变量建立包含 2 个

隐藏层、每层 10 个神经元的全连接神经网络模型,各血常规指标为输入层, α 、 β 地贫为输出层,训练集结果 α 地贫预测值为 90.845%、 β 地贫预测值为 91.729%。预测结果没有达到 100% 的原因可能为模型特征不够精确所致,而 α 地贫预测值较 β 地贫预测值低,可能是 β 地贫贫血表现较 α 地贫更严重,导致模型算法出现偏倚。为提高模型预测值,运用连接权重算法进行 α 、 β 地贫各变量重要性分析,绝对值越高代表重要性越高,正负代表正相关和负相关,得出 MCV、RDW-CV、MCH、RBC、RDW-SD、Hb 6 个重要性最高的变量,与随机森林模型、Logistic 模型数据分析结果一致。随后基于变量权重,对 α 、 β 地贫神经网络预测模型进行优化,将以上 6 个变量作为输入层,设置 2 个隐藏层、每层 5 个神经元, α 、 β 地贫作为输出层,输入验证集数据, α 地贫、 β 地贫预测值均为 100%。见图 3。



注:地贫预测模型可视化,总分对应线性预测值(Linear Predictor 分数) >0 提示地贫阳性。

图 2 地贫特征变量预测模型的建立



注:A 为训练集模型;B 为模型预测值;C 为连接权重法进行变量重要性分析(绝对值越高,重要性越高);D 为优化后模型;E 为优化后模型预测值。

图 3 α 、 β 地贫预测效果分析

3 讨 论

随着人工智能技术的飞速发展,运用神经网络等技术辅助诊断疾病成为国内外研究热点。与人类诊断相比,基于计算机的诊断方法可以产生更准确的输出结果和更高的效率,有效减少误诊率,从而以更少的成本实现更有效的疾病诊断^[3,15]。通过大数据和临床信息的诊断,神经网络极大地提高了临床诊断结果的准确性和检验效率,促进检验医学向精准医学、智能检验快速发展^[16-17]。如果将神经网络运用于地贫的早期诊断,用普及程度较高的血常规就能预测严重程度较高的地贫型别,降低地贫基因鉴定的高额费用及等待时间,从而及早采取治疗措施,并对地贫孕妇开展基因检测及遗传咨询,这对降低中重型地贫患儿出生率,以及西南、西北等中低收入地区的人群健康普查有重大意义。

本研究对 9 652 例孕妇血常规指标和地贫基因检测结果进行分析,检出 α 地贫阳性率为 5.90%, β 地贫阳性率为 2.77%,地贫型别分布与杨嫒等^[18] 研究相似。本研究通过对地贫危险因素进行分析发现,RBC、RDW-CV、中孕期、晚孕期为地贫的危险因素。随孕期进展,地贫孕妇贫血症状加剧,RBC、RDW-CV 增大则为机体造血代偿性增生,导致红细胞数量增多且大小分布不均。Hb、HCT、MCV、MCH、MCHC、RDW-SD、早孕期为地贫的保护性因素,其结果与地贫呈负相关,即其值越高,地贫发生的概率越低,与既往报道一致^[1,2-4]。随后,本研究采用随机森林模型筛选出预测地贫重要性较高的几个系数:MCV、MCH、RDW-CV、MCHC、RBC。这与文献^[19-20]的研究结果相近。采用 AUC 评价模型筛选效果,MCV、MCH 分数最高,且试验组对照组结果一致。综合试验组和对照组,MCH、MCV 的预测价值最高,与地贫专家共识^[21]将 $MCV < 82$ fL、 $MCH < 27$ pg 作为地贫的血常规筛查指标吻合。

随机森林是一种集成学习算法,可减少过拟合的风险,提高了模型的泛化能力,更接近实际情况^[22]。本研究建立神经网络模型,输入层 1 层,11 个神经元数量,将所有指标全部纳入;地贫的血常规指标有多个特征,且不能进行简单的线性分离,故设置隐藏层 1 层,神经元数量为 10,尽可能提取到各输入变量的特征;输出层 1 层,2 个节点分别为 α 、 β 。与其他运用 AI 学习工具预测地贫的研究相比,研究者创新性地加入了基于反向传播(BP)神经网络的 RPROP 算法,运用带有权重回溯的弹性反向传播更新权重,以降低预测值与实际值的误差。输入训练集,得到 α 地贫预测值为 90.845%、 β 地贫预测值为 91.729%。怀疑初始模型由于输入层神经元数量和隐藏层节点数量过多,且输入层神经元特征不够代表性,导致模型过拟合,预

测效果反而下降。因此,对模型进行优化,减少输入层神经元数量,将重要性最高的 6 个变量作为输入节点,隐藏层数量降为 5,输入验证集进行验证, α 地贫、 β 地贫预测值均提升为 100%。

有研究证明,随机森林算法在多种疾病的诊断和预后预测中具有较高的效率,且已有研究将随机森林算法运用于 α 地贫中^[23-24],其预测效率与本研究模型相当。神经网络已被证明在包括测序在内的多种数据中可以被用于疾病的诊断^[25],且在地贫的预测中有较高的效率^[26],但目前为止仍鲜有对 α 地贫和 β 地贫进行区分的模型报道。本研究模型集合了两种有效的算法,不仅利用较为容易获取的临床数据通过随机森林对地贫进行了预测,且通过神经网络构建模型对 α 地贫和 β 地贫进行了区分。不过,由于 α 地贫主要数据为 $-\alpha^{3.7}$ 、 $-\alpha^{SEA}$ 、 $-\alpha^{4.2}$, β 地贫主要数据为 CD17(A-T)、CD41-42(-TCTT)、IVS-II-654(C-T),实际预测结果会向标本量较大的地贫基因型倾斜。各地贫基因型的数据量与地贫阴性数据量差异较大,且孕妇随着孕周增大,生理性贫血加剧,阴性数据中,贫血数据的比重增大,对预测模型有一定误导,导致具体地贫基因型预测效果不佳。此外,由于神经网络模型具有一定的随机性,以及重庆市地贫分布的局限性,对地贫孕妇的预测效果有待扩大数据量及分布范围进一步研究。同时,由于缺铁性贫血同为小细胞低色素性贫血,血常规指标与地贫有部分重叠,对于地贫和缺铁贫的鉴别将作为下一步研究重点。

综上所述,本研究建立神经网络模型对 9 652 例孕妇进行地贫基因预测,将孕妇血常规指标作为模型变量,筛选出 MCV、RDW-CV、MCH、RBC、RDW-SD、Hb 6 个重要性较高的指标,并通过将以上 6 项指标作为变量,得出 α 地贫、 β 地贫预测值均为 100%。本研究结果可为地贫的早期筛查提供新的思路。

参考文献

- [1] KATTAMIS A, KWIATKOWSKI J L, AYDINOK Y. Thalassaemia[J]. Lancet, 2022, 399(10343): 2310-2324.
- [2] TAHER A T, WEATHERALL D J, CAPPELLINI M D. Thalassaemia[J]. Lancet, 2018, 391(10116): 155-167.
- [3] MO D, ZHENG Q, XIAO B, et al. Predicting thalassemia using deep neural network based on red blood cell indices [J]. Clin Chim Acta, 2023, 543: 117329.
- [4] WANG W D, HU F, ZHOU D H, et al. Thalassaemia in China[J]. Blood Rev, 2023, 60: 101074.
- [5] ECKRICH M J, FRANGOUL H. Gene editing for sickle cell disease and transfusion dependent thalassemias: a cure within reach[J]. Semin Hematol, 2023, 60(1): 3-9.
- [6] EL AFIFI A M, SAEED A M, FEKRY G H, et al. Suc-

- cessful management of severe hyperhaemolysis with combined tocilizumab and rituximab in non-transfusion-dependent thalassaemia; a case report [J]. *Transfus Med Hemother* [J], 2022, 50(1):66-70.
- [7] VIPRAKASIT V, EKWATTANAKIT S. Clinical classification, screening and diagnosis for thalassemia [J]. *Hematol Oncol Clin North Am*, 2018, 32(2):193-211.
- [8] ACHOUR A, KOOPMANN T T, BAAS F, et al. The evolving role of next-generation sequencing in screening and diagnosis of hemoglobinopathies [J]. *Front Physiol*, 2021, 12:686689.
- [9] NARANG V, JAIN A, GROVER S, et al. Prevalence and spectrum of haemoglobinopathies in females of reproductive age group-A first tertiary care center experience in Punjab, North India [J]. *Indian J Pathol Microbiol*, 2023, 66(3):564-567.
- [10] ALMOMANI A A, SHRAIM A S, ATOOM A M, et al. Evaluation of the validity of the pre-marriage mean corpuscular volume value as a predictive test for β -thalassaemia carrier status [J]. *J Med Biochem*, 2023, 42(2):195-205.
- [11] THILAKARATHNE S, JAYAWEERA U P, PREMAWA RDHENA A. Unresolved laboratory issues of the heterozygous state of β -thalassaemia; a literature review [J]. [2023-06-10]. <https://pubmed.ncbi.nlm.nih.gov/37259577/>.
- [12] SOONTORN PANAWET C, SINGHA K, SRIVORAKUN H, et al. Molecular basis of a high Hb A2/Hb F β -thalassaemia trait; a retrospective analysis, genotype-phenotype interaction, diagnostic implication, and identification of a novel interaction with α -globin gene triplication [J]. *PeerJ*, 2023, 11:e15308.
- [13] JOOSSE H J, VAN OIRSCHOT B A, KOOIJMANS S A A, et al. In-vitro and in-silico evidence for oxidative stress as drivers for RDW [J]. *Sci Rep*, 2023, 13(1):9223.
- [14] WACKA E, WAWRZYNIAK-GRAMACKA E, TYLUTKA A, et al. The role of inflammation in age-associated changes in red blood system [J]. *Int J Mol Sci*, 2023, 24(10):8944.
- [15] AZLAN N O, SHAHARUL A A M A, SALWA N D, et al. Optimization of identifying insulinaemic pharmacokinetic parameters using artificial neural network [J]. *Computer methods and programs in biomedicine*, 2023, 236.
- [16] 余恋雨, 陈鸣. 人工智能在检验领域的新发展新方向 [J]. *中华检验医学杂志*, 2021, 44(10):887-891.
- [17] RABBANI N, KIM G Y E, SUAREZ C J, et al. Applications of machine learning in routine laboratory medicine: current state and future directions [J]. *Clin Biochem*, 2022, 103:1-7.
- [18] 杨嫻, 朱丽丹, 张颖, 等. 重庆地区 108 140 例疑似贫血患者地中海贫血筛查及检出基因型分析 [J]. *第三军医大学学报*, 2020, 42(17):1750-1756.
- [19] PAN Y, CHEN M, ZHANG Y, et al. Analysis of genotype-phenotype correlation in patients with α -thalassaemia from Fujian Province, Southeastern China [J]. *J Clin Lab Anal*, 2022, 36(10):e24696.
- [20] HERNANINGSIH Y, SYAFITRI Y, INDRASARI Y N, et al. Analysis of common beta-thalassaemia (β -thalassaemia) mutations in east java, indonesia [J]. *Front Pediatr*, 2022, 10:925599.
- [21] 地中海贫血妊娠期管理专家共识 [J]. *中华围产医学杂志*, 2020, 23(9):577-584.
- [22] GHOSH D, CABRERA J. Enriched random forest for high dimensional genomic data [J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2022, 19(5):2817-2828.
- [23] FENG P, LI Y, LIAO Z, et al. An online alpha-thalassaemia carrier discrimination model based on random forest and red blood cell parameters for low HbA2 cases [J]. *Clin Chim Acta*, 2022, 525:1-5.
- [24] YAO Z, ZHANG W, SONG P, et al. Deepformer: a hybrid network based on convolutional neural network and flow-attention mechanism for identifying the function of DNA sequences [J]. *Brief Bioinform*, 2023, 24(2):bbad095.
- [25] MO D, ZHENG Q, XIAO B, et al. Predicting thalassaemia using deep neural network based on red blood cell indices [J]. *Clin Chim Acta*, 2023, 543:117329.
- [26] XU M, LIN G, DONG Z, et al. Logistic-Nomogram model based on red blood cell parameters to differentiate thalassaemia trait and iron deficiency anemia in southern region of Fujian Province, China [J]. *J Clin Lab Anal*, 2023, 37(11/12):e24940.

(收稿日期:2023-06-29 修回日期:2023-07-16)